Coherence Requirements for Belief

Review & Stage-Setting

- Premise (1) seems to (inevitably) smuggle-in some probabilistic assumptions (via propriety/Pr-admissibility). In this sense, such arguments seem “circular” (to some extent).
- From now on, I will simply assume that $s(\cdot,E)$ is always a probability function, i.e., that (PRE) is correct.
- However, the argument above is still interesting because it introduces the following key theoretical notion.

**Evidential Consistency** (EC). A judgment set is *evidentially consistent* just in case there exists *some* body of total evidence $E$ which supports each of its members.

- (EC) allows us to restate (3) & (4) in the above argument as:
  3. If $b(\cdot)$ is (weakly) dominated in $2_B$-accuracy, then $b(\cdot)$ is *evidentially inconsistent*.
  4. $\therefore$ if (1) is true, then $b(\cdot)$ is evidently consistent *only if* $b(\cdot)$ — and $s(\cdot,E)$ — is probabilistic.

- (EC) will play a key role in the application to full belief.

Lecture #2: Full Belief

Kenny Easwaran 1  Branden Fitelson 2

1Philosophy @ USC
easwaran@usc.edu
2Philosophy & RuCCS @ Rutgers
&
MCM & LMU
branden@fitelson.org

Premise (1) seems to (inevitably) smuggle-in some probabilistic assumptions (via propriety/Pr-admissibility). In this sense, such arguments seem “circular” (to some extent).

From now on, I will simply assume that $s(\cdot,E)$ is always a probability function, i.e., that (PRE) is correct.

However, the argument above is still interesting because it introduces the following key theoretical notion.

**Evidential Consistency** (EC). A judgment set is *evidentially consistent* just in case there exists *some* body of total evidence $E$ which supports each of its members.

- (EC) allows us to restate (3) & (4) in the above argument as:
  3. If $b(\cdot)$ is (weakly) dominated in $2_B$-accuracy, then $b(\cdot)$ is *evidentially inconsistent*.
  4. $\therefore$ if (1) is true, then $b(\cdot)$ is evidently consistent *only if* $b(\cdot)$ — and $s(\cdot,E)$ — is probabilistic.

- (EC) will play a key role in the application to full belief.

Review from last time. We’re after *formal, synchronic, epistemic, coherence requirements* (of ideal rationality).

- Our strategy is to apply a Joyce-style accuracy-dominance approach to three different types of judgment sets.
- Last time, we examined Joycean arguments for *probabilism* as a CR for credences. We prefer to think of these as arguments that *both* $b(\cdot)$ and $s(\cdot,E)$ are probabilistic.

1. The overall degree of inaccuracy of $b(\cdot)$ — at $w$ — is (and should be) measured by some function $2_B(b(\cdot), v_w(\cdot))$, which is *proper* (or, at least, *probabilistically admissible*). &
2. If (1) is true, then $b(\cdot)$ is non-(weakly)-dominated in $2_B$-accuracy iff $b(\cdot)$ is *probabilistic*.
3. If $b(\cdot)$ is (weakly) dominated in $2_B$-accuracy, then — *no matter what the total evidence* $E$ — is $b(\cdot) \neq s(\cdot,E)$.
4. $\therefore$ if (1) is true, then $b(\cdot) = s(\cdot,E)$ — for *some* body of total evidence $E$ — *only if* $b(\cdot)$ — and $\therefore s(\cdot,E)$ — is probabilistic.

Here is a — perhaps the — “paradigm” CR [28, 30, 27, 22].

- **The Consistency Requirement for Belief Sets** (CB). All belief sets should be *logically* (viz., *alethically*) consistent.
- (CB) follows from the following (narrow-scope) norm:
  - **The Truth Norm for Belief** (TB). All (individual) beliefs should (alethically–ideally) be *true*.
- Alethic norms [(CB)/(TB)] can conflict with evidential norms.
  - **The Evidential Norm for Belief** (EB). All (individual) beliefs should (evidentially–ideally) be *supported by the evidence*.
  - In some cases (*e.g.*, preface cases), agents satisfy (EB) while violating (CB) — this generates an alethic/evidential *conflict*.
  - Such alethic/evidential conflicts needn’t give rise to states that receive an (overall) evaluation as *irrational* (nor must they inevitably give rise to rational *dilemmas*) [6, 24, 14, 23].
  - We’ll refer to the claim that there exist *some* such cases as *the datum*. Foley’s [14] explanation of *the datum* is helpful.
“...if the avoidance of recognizable inconsistency were an absolute prerequisite of rational belief, we could not rationally believe each member of a set of propositions and also rationally believe of this set that at least one of its members is false. But this in turn pressures us to be unduly cautious. It pressures us to believe only those propositions that are certain or at least close to certain for us, since otherwise we are likely to have reasons to believe that at least one of these propositions is false. At first glance, the requirement that we avoid recognizable inconsistency seems little enough to ask in the name of rationality. It asks only that we avoid certain error. It turns out, however, that this is far too much to ask.”

- We will not argue for the datum here. We think Foley [14], Christensen [6], Kolodny [24], and others have made a compelling case for it. Today, it is our point of departure. [But, we do have a new Preface case that we think is definitive.]

Some philosophers construe the datum as reason to believe that \((\ast)\) there are no coherence requirements for full belief.

- Christensen [6] thinks (a) credences do have coherence requirements (probabilism); \((\ast)\) full beliefs do not; (b) what seem to be CRs for full belief can be explained via (a).

- Kolodny [24] agrees with \((\ast)\), but he disagrees with (a) and (b). He thinks (c) full belief is explanatorily indispensable; (d) there are no coherence requirements for any judgments; (e) what seem to be CRs for full belief can be explained via (EB).

- Christensen & Kolodny agree — trivially, via \((\ast)\) — that:

\((\dagger)\) If there are any coherence requirements for full belief, then (CB) is a coherence requirement for full belief.

- We [2, 11] agree with Christensen on (a) and Kolodny on (c), but we disagree with them on \((\ast)\), (d), (e), and \((\dagger)\). We’ll explain how to ground “conflict-proof” CRs for full belief, by analogy with Joyce’s [21, 19] argument(s) for probabilism.

We begin with some background assumptions/notation.

- \(B(p) \equiv S\) believes that \(p\). \(D(p) \equiv S\) disbelieves that \(p\).

- \(S\) makes judgments regarding propositions in a (finite) agenda \((A)\) of (classical, possible-worlds) propositions. We’ll use “\(B\)” to denote the set of \(S\)’s judgments on \(A\).

- We’re only evaluating explicit judgments (on \(A\)) — we assume nothing about off-\(A\) (“implicit”) commitments.

- We’ll assume the following about \(B/D\) on \(A\). The first assumption is integral to the framework. The second two assumptions are made for simplicity (and could be relaxed).

  - Accuracy conditions. \(B(p) \mid D(p))\) is accurate iff \(p\) is \(T [F]\).

  - Incompatibility. \(B(p) \Rightarrow \neg D(p)\).

  - Opinionation. \(B(p) \lor D(p)\).


- We assume belief/world independence — until slide (17).
Now, we can explain how our new CRs were discovered, by analogy with Joyce’s [21, 19] argument(s) for probabilism.

Both arguments can be seen as involving three key steps.

**Step 1:** Define \( \hat{B}_w \) — the vindicated (viz., aethically ideal or perfectly accurate) judgment set (on \( A \)), at world \( w \).
- \( \hat{B}_w \) contains \( B(p) \) iff \( p \) is true (false) at \( w \).
- Heuristically, we can think of \( \hat{B}_w \) as the set of judgments that an omniscient agent would have (on \( A \), at \( w \)).

**Step 2:** Define \( d(B, B_w) \) — a measure of distance between \( B \) and \( B_w \). That is, a measure of \( B \)’s overall inaccuracy (at \( w \)).
- \( d(B, B_w) \triangleq \) the number of inaccurate judgments in \( B \) at \( w \).
- *Hamming distance* [8] between the binary vectors \( B, \hat{B}_w \).

**Step 3:** Adopt a fundamental epistemic principle, which uses \( d(B, B_w) \) to ground a coherence requirement for \( B \).
- This last step is the philosophically crucial one...

Ideally, we want a coherence requirement that [like (CB)] can be motivated by considerations of *accuracy* (viz., a CR that is *entailed by* alethic requirements such as TB/CB/PV).

But, in light of (e.g.) preface cases, we also want a CR that is *weaker* than (CB). More precisely, we want a CR that is weaker than (CB) *in such a way that it is also entailed by* (EB).

We can show that our new CRs [e.g., (WADA)] fit the bill, *if* we assume the following “probabilistic-evidentialist” necessary condition for the satisfaction of (EB).

**Necessary Condition for Satisfying (EB).** \( B \) satisfies (EB), i.e., all judgments in \( B \) are *supported by the evidence*, only if:

\[
(R) \quad \text{There exists some Pr-function that probabilifies (i.e., assigns Pr greater than } 1/2 \text{ to) each belief in } B \text{ and dis-probabilifies (i.e., assigns Pr less than } 1/2 \text{ to) each disbelief in } B. \]

“Probabilistic-evidentialists” will disagree about which Pr(·) undergirds (EB) [5, 32, 15, 20]; but, they agree on (EB) ⇒ \( (R) \). In fact, Pr-evidentialists will agree that (EB) ⇒ (EC) ⇒ \( (R) \).

### Possible Vindication (PV)
There exists some possible world \( w \) at which all of the judgments in \( B \) are accurate. Or, to put this more formally, in terms of \( d \): \( \exists w \). \( d(B, B_w) = 0 \).

Possible vindication is *one* way we could go here. But, our framework is much more general than the classical one. It allows for many other choices of fundamental principle.

Like Joyce [21, 19] — who makes the analogous move with credences, to ground probabilism — we retreat from (PV) to the weaker: avoidance of (weak) dominance in \( d(B, B_w) \).

#### Weak Accuracy-Dominance Avoidance (WADA)

- There does not exist an alternative belief set \( B’ \) such that:
  1. \( \forall w \). \( d(B’, \hat{B}_w) \leq d(B, B_w) \), and
  2. \( \exists w \). \( d(B’, \hat{B}_w) < d(B, B_w) \).
- Completing Step 3 in this way reveals new CRs for \( B \)...

#### Logical Relationships

| (TB) | (EB) |
| (CB)/(PV) | (EC) |

\[ \Rightarrow \] (WADA)

See slide #20 for a bigger map w/11 requirements/norms.
The key to our central theorem that \((R) \Rightarrow (WADA)\) is that our inaccuracy measure \(d(B, B_w)\) is \textit{evidential proper}.

**Definition (Evidential Propriety)**

Suppose a judgment set \(J\) of type \(J\) is supported by the evidence. That is, suppose there exists some evidential probability function \(\Pr(\cdot)\) which represents \(J\) (in the appropriate sense of "represents" for sets of type \(J\)). If this is sufficient to ensure that \(J\) minimizes expected inaccuracy (relative to \(\Pr\)), according to the measure of inaccuracy \(I(J, J_w)\), then we will say that the measure \(I\) is \textit{evidentially proper}.

If an inaccuracy measure is \textit{evidentially improper}, then some probabilistically representable judgment sets will be \textit{ruled out} as \textit{irrational} via accuracy-dominance (WADA).

This would engender a \textit{conflict} between alethic and evidential requirements for judgment, which is exactly what coherence requirements are \textit{not} supposed to do.

In our book [12], evidential propriety plays a central role.

There are many advantages to adopting \((R)\), rather than (WADA), as our (ultimate) CR for full belief. Here are a few:

- First, (WADA) is (intuitively) \textit{too weak} to serve as our (ultimate) CR — \(\{B(p), B(\neg p)\}\) may be \textit{non-dominated}, as the following table reveals (\textit{ditto} for \(\{D(p), D(\neg p)\}\)).

<table>
<thead>
<tr>
<th>(w_1)</th>
<th>(P)</th>
<th>(\neg P)</th>
<th>(B(P))</th>
<th>(B(\neg p))</th>
<th>(D(\neg p))</th>
<th>(D(p))</th>
<th>(D(\neg p))</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

- \((R) \Rightarrow (NCP)\) \(D(p) \equiv B(\neg p)\), which rules-out \(\{B(p), B(\neg p)\}\).

- \((R)\) is strictly stronger than (WADA) + (NCP). Indeed, we conjecture that \((R)\) is the strongest CR (uncontroversially) entailed by both alethic and evidential considerations.

- (WADA) only makes sense for \textit{finite} agendas, whereas \((R)\) is potentially applicable to \textit{infinite} agendas (if there be such).

- \((R)\) entails (WADA\(_d\)), for \textit{any additive inaccuracy measure} \(d\).

In this sense, \((R)\) is \textit{robust} across choices of \(d\).

Proof of the central result that \((R) \Rightarrow (WADA)\).

Let \(\Pr\) be a probability function that represents \(B\) in sense of \((R)\). Consider the \textit{expected} inaccuracy of a belief set — the sum of \(\Pr(w) \cdot d(B, B_w)\). Since \(d(B, B_w)\) is a sum of components for each proposition \(1\) if \(B\) disagrees with \(w\) on the proposition and \(0\) if they agree), and since expectations are linear, the expected inaccuracy is the sum of the expectation of these components. The expectation of the component for disbelieving \(p\) is \(\Pr(p)\) while the expectation of the component for believing \(p\) is \(1 - \Pr(p)\). Thus, if \(\Pr(p) > 1/2\) then believing \(p\) is the attitude that uniquely minimizes the expectation, while if \(\Pr(p) < 1/2\) then disbelieving \(p\) is the attitude that uniquely minimizes the expectation. Thus, since \(\Pr\) represents \(B\), this means that \(B\) has strictly lower expected inaccuracy than any other belief set with respect to \(\Pr\). Suppose, for \textit{reductio}, that some \(B'\) (weakly) dominates \(B\). Then, \(B'\) must be no farther from vindication than \(B\) in any world, and thus \(B'\) must have had expected inaccuracy no greater than that of \(B\). But \(B\) has strictly lower expected inaccuracy than any other belief set. Contradiction. \(\therefore\) \(B\) must be non-dominated.

It is useful to draw an analogy between the norms and requirements we’ve been discussing, and principles in rational choice theory. The Decision-Theoretic Analogy.

<table>
<thead>
<tr>
<th>Epistemic Principle</th>
<th>Analogous Decision-Theoretic Principle</th>
</tr>
</thead>
<tbody>
<tr>
<td>(TB)</td>
<td>(AMU) Do (\phi) if (\phi) maximizes utility in the \textit{actual} world.</td>
</tr>
<tr>
<td>(CB)</td>
<td>(PMU) Do (\phi) if (\phi) maximizes (u) in \textit{some possible} world.</td>
</tr>
<tr>
<td>(R)</td>
<td>(MEU) Do (\phi) if (\phi) maximizes EU (relative to some (\Pr)).</td>
</tr>
<tr>
<td>(WADA)</td>
<td>(WDOM) Do (\phi) if (\phi) is \textit{not weakly dominated} in utility.</td>
</tr>
<tr>
<td>(SADA)</td>
<td>(SDOM) Do (\phi) if (\phi) is \textit{not strictly dominated} in utility.</td>
</tr>
</tbody>
</table>

Like (TB), (AMU) is \textit{not a requirement of rationality}; and, like (CB), (PMU) isn’t a rational requirement either. Moreover, also like (CB), seeing this requires “paradoxical” cases [25].

As Foley \textit{(op. cit.)} explains, (CB) is \textit{too demanding}. But, (R) and (WADA) are \textit{not} — they do not “pressure us to believe only those propositions that are (close to) certain for us.”

There are analogous examples for full belief. Consider:

- One can argue (Caie-style) that the only non-dominated (opinionated) belief sets on \( \{P, \neg P\} \) are \( \{B(P), B(\neg P)\} \) and \( \{D(P), D(\neg P)\} \), which are both ruled-out by \( (R) \).

The “\( \times \)’s indicate that these worlds are ruled-out \( (a \text{ } \text{priori}) \) by the definition of \( P \). As such, the only non-dominated belief sets seem to be \( \{B(P), B(\neg P)\} \) and \( \{D(P), D(\neg P)\} \).

If this Caie-style reasoning is correct, then it shows that some of our assumptions must go. But, which one(s)?

We have the following four facts regarding \( B_1 \) & \( B_2 \):

1. \( B_1 \) is weakly dominated in inaccuracy by \( B_2 \) (this is easily verified by simple counting). Thus, \( B_1 \) violates (WADA).
2. \( B_1 \) is not strictly dominated in inaccuracy by any belief set over \( B \) (this can be verified via exhaustive search on the set of all belief sets over \( B \)). Thus, \( B_1 \) satisfies (SADA).
3. \( B_2 \) is not weakly dominated (in inaccuracy) by any belief set over \( B \) (this can be verified via exhaustive search on the set of all belief sets over \( B \)). Thus, \( B_2 \) satisfies (WADA).
4. \( B_2 \) is not represented (in the sense of Definition 2) by any probability function on \( B \), since the set \( B_2 \) contains two contradictory pairs: \{\( D(Y), D(\neg Y) \}\}, \{\( D(X \equiv Y), D(\neg(X \equiv Y)) \}\). Therefore, \( B_2 \) violates (R).

Here is what the logical relations look like, among all of the 11 requirements & norms for (opinionated) full belief:

- (TB) S ought believe \( p \) just in case \( p \) is true.
- (PV) \( (\exists w)[d(B, B_w) = 0] \). \( (B \text{ is logically/alethically consistent.}) \)
- (SADA) \( \exists B' \) such that: \( (\forall w)[d(B', B_w) < d(B, B_w)] \).
- (NW2S) \( \exists \beta \subseteq B \text{ s.t.: } (\forall w)[ > \frac{1}{2} \text{ of the members of } \beta \text{ are inaccurate at } w] \).
- (Rr) \( \exists \text{ a probability function Pr(\cdot) such that, } \forall p \in A: \)
  \( B(p) \text{ iff } \text{Pr}(p) > r, \text{ and } D(p) \text{ iff } \text{Pr}(p) < 1 - r \).
- (EB) S ought believe \( p \) just in case \( p \) is supported by S’s evidence. Note: this assumes only \( (\exists \text{Pr})(\forall p) [\text{Pr}(p) > \frac{1}{2} \text{ iff } B(p)] \).
- (NWS) \( \exists \beta \subseteq B \text{ s.t.: } (\forall w)[ > \frac{1}{2} \text{ of the members of } \beta \text{ are inaccurate at } w] \)
- (WADA) \( \exists B' \text{ s.t.: } (\forall w)[d(B', B_w) \leq d(B, B_w)] \& (\exists w)[d(B', B_w) < d(B, B_w)] \).
- (NW1S) \( \exists \beta \subseteq B \text{ s.t.: } (\forall w)[ > \frac{1}{2} \text{ of the members of } \beta \text{ are inaccurate at } w] \).
- (NCP) S disbelieves \( p \) iff S believes \( \neg p \) [i.e., \( D(p) \equiv B(\neg p) \)].
Proof of the claim that (NWS) ⇔ (WADA).

(⇒) We’ll prove the contrapositive. Suppose that some S ⊆ B is a witnessing set. Let B′ agree with B on all judgments outside S and disagree with B on all judgments in S. By the definition of a witnessing set, B′ weakly dominates B in inaccuracy [d(B, Bw)].

(⇐) [Contrapositive again.] Suppose B is dominated, i.e., that there is some B′ that weakly dominates B in inaccuracy [d(B, Bw)]. Let S ⊆ B be the set of judgments on which B and B′ disagree. Then, S is a witnessing set.

A similar proof can be given for: (NW1S) ⇔ (SADA).

We also know that (R) ⇒ (NW2S). See next slide for a proof.

The converse remains an open question. That is:

(O) (NW2S) ⇔ (R).

We do know that (R) is strictly stronger than the conjunction (WADA) & (NCP). See the slide #23 for a proof.

---

**Theorem**

(NDB & NCP) /arrowdblrightnot(R). [In other words, (WADA & NCP) /arrowdblrightnot(R).]

**Proof.**

Let there be six possible worlds, w1, w2, w3, w4, w5, w6. And, let ∙A ∋ {p1, p2, p3, p4}, where the pi are defined as follows.

\[
p_1 ∋ \{w_1, w_2, w_3\} \quad p_2 ∋ \{w_1, w_4, w_5\} \\
p_3 ∋ \{w_2, w_4, w_6\} \quad p_4 ∋ \{w_3, w_5, w_6\}
\]

Let B ∋ {B(p1), B(p2), B(p3), B(p4)}. B is a witnessing2 set, since, in every w1, exactly half of the beliefs in B are accurate. So, by (R) ⇒ (NW2S), B violates (R). But, B satisfies (NDB), since every belief set on ∙A has an expected inaccuracy of 2, relative to the uniform Pr-distribution, which implies that no belief set on ∙A dominates any other belief set on ∙A. Finally, B satisfies (NCP), since every pair of beliefs in B is consistent.

---

**Parametric Family of Requirements Between (R) and (CB)**

(Rr) There is a probability function Pr such that, for all p ∋ A:

(i) B contains B(p) iff Pr(p) > r, and

(ii) B contains D(p) iff Pr(p) < 1 − r,

where r ∈ [1/2, 1).

Let ∙Bn denote the class of minimal inconsistent belief sets of size n — each member of ∙Bn is an inconsistent judgment set of size n containing no inconsistent proper subset.

Let Bn be a member of ∙Bn, i.e., Bn consists of n propositions, there is no world in which all of these n propositions are true, but for each proper subset B ⊂ Bn there is a world in which all members of B are true.

**Theorem**

For all n ≥ 2, if r ≥ \(\frac{n-1}{n}\), then (Rr) rules out each member of ∙Bn, while if r < \(\frac{n-1}{n}\), then (Rr) rules out no member of ∙Bn.
We (along with Rachael Briggs and Fabrizio Cariani) [2] are investigating various applications of this new approach.

One interesting application is to judgment aggregation. E.g.,

- Majority rule aggregations of the judgments of a bunch of agents — each of whom satisfy (PV) — need not satisfy (PV).

- Q: does majority rule preserve our notion of coherence, viz., (WADA) preserved by MR? A: yes (on simple, atomic + truth-functional agendas), but not on all possible agendas.

- There are (not merely atomic + truth-functional) agendas A and sets of judges J (|A| ≥ 5, |J| ≥ 5) that (severally) satisfy (WADA), while their majority profile violates (WADA).

- But, if a set of judges is (severally) consistent [i.e., satisfy (PV)], then their majority profile must satisfy (WADA).

Recipe. Wherever B-consistency runs into paradox, substitute coherence (in our sense), and see what happens.

Kenny has written a paper [10] that explains how to relax the assumption of Opinionation in our framework.

Our approach is equivalent to assigning (in)accurate judgments a score of (−1) + 1, and calculating the total score of B (at w) as the sum of the scores of all p ∈ A.

Kenny’s Generalizations: (a) allow scores of −w and +r, where w ≥ r > 0, and (b) allow S to suspend on p [S(p)], where all suspensions are given a neutral score of zero.

This generalization of our framework leads to an elegant analogue of our central Theorem that (R) entails (WADA).

**Theorem.** An agent S will avoid (strict) dominance in total score if their belief set B can be represented as follows:

\[(R) \text{ There exists a probability function } Pr(\cdot) \text{ such that, } \forall p \in A:\]

\[B(p) \iff Pr(p) > \frac{w}{r+w},\]

\[D(p) \iff Pr(p) < 1 - \frac{w}{r+w},\]

\[S(p) \iff Pr(p) \in \left[1 - \frac{w}{r+w}, \frac{w}{r+w}\right].\]
<table>
<thead>
<tr>
<th>Refs</th>
<th>Coherence Requirements for Belief</th>
<th>Extras</th>
</tr>
</thead>
</table>